



Title	Influence maximization in noncooperative social networks
Author(s)	Yang, Y; Li, VOK; Xu, K
Citation	The 2012 IEEE Global Communications Conference (GLOBECOM), Anaheim, CA., 3-7 December 2012. In Globecom. IEEE Conference and Exhibition, 2012, p. 2834-2839
Issued Date	2012
URL	http://hdl.handle.net/10722/191614
Rights	Globecom. IEEE Conference and Exhibition. Copyright © IEEE.

Influence Maximization in Noncooperative Social Networks

Yile Yang, Victor O.K. Li, Kuang Xu
University of Hong Kong, Pokfulam, Hong Kong, China

Abstract—In this paper, we consider the problem of maximizing information propagation with noncooperative nodes in social networks. We generalize the linear threshold model to take node noncooperation into consideration and provide a provable approximation guarantees for the noncooperative influence maximization problem. We propose an analytical model based on the generalized maximum flow problem to characterize the noncooperative behavior of an individual node in maximizing influence. Based on this, we develop a new seed node selection strategy, under the linear threshold model, to account for user noncooperativeness. Extensive simulations on large collaboration networks show that our proposed flow-based strategy outperforms the weighted degree scheme under various noncooperative scenarios. The evaluation also validates the importance of cooperation and incentives in maximizing influence.

I. INTRODUCTION

In online advertising campaigns, identifying pilot customers is important in order to maximize the advertisement influence among the whole target audience. The prevalence of online social networking sites such as Facebook and Twitter [1] offers the possibility of mining for necessary information. Normally, initially activated seed nodes are more cooperative to forward the product information because being selected as early adopters makes them feel “special”. Also, incentives offered by companies may foster initial users’ sense of duty to recommend the products to their neighbors in the social network [2]. However, other ordinary (i.e. non-seed) users may not pass along the information because of the costs of resources (time, personal information, etc.) and lack of incentive. Existing work focuses on the formulation of related influence propagation models [3] and optimization problems [4] [5], as well as on seed node selection algorithms which can achieve the near-optimal performance [6], but they do not study the heterogeneity of cooperativeness between the seed and the ordinary nodes in the influence propagation process.

In this work, firstly, we generalize the standard linear threshold model (LTM) to take node noncooperation into consideration, and provide a provable approximation guarantees for the corresponding model. In a noncooperative influence diffusion model we define a node in the system to be noncooperative if it chooses to reserve some of its resources during the influence propagation process. Specifically, in noncooperative LTM it means that a noncooperative active node will not use its link’s full capacity, which is the weight of edges in the social graph to influence its inactive neighbors. Secondly, we propose an analytical model, based on the generalized maximum flow problem, that characterizes the behavior of nodes in

influence maximization problem. Drawing on maximum flow, we are able to determine the maximal amount of influence between pairs of nodes on the basis of all independent paths in the graph. We believe this more accurately depicts the influence propagation process in social networks. In addition, we propose a new seed node selection strategy for LTM to account for node noncooperativeness, under the example of the recently proposed cluster-based heuristic strategy [7]. We also conducted extensive simulations on large collaboration networks, and the result not only shows that node cooperation as well as incentives play fundamental roles in determining performance of influence maximization algorithms, but also proves that our proposed strategy is more robust when nodes are not cooperative.

The remainder of this paper is structured as follows. Related work is described in Section II. We present the problem formulation and diffusion model in Section III. In Section IV we describe the generalized noncooperative LTM and provide some nice properties of the model, followed by the description of the noncooperative influence maximization problem and the proposed flow-based centrality measure. The evaluation is presented in Section V. Finally, we conclude this study with suggestions for future work in Section VI.

II. RELATED WORK

Network flow problems [8], especially the maximum flow problem, have been important problems in networking research. They have recently found applications in online reputation systems [9] and P2P incentive paradigm design [10]. In [11], the authors introduce a modified betweenness-based centrality measure which is based on the concept of network flows. While the traditional betweenness¹ metric measures the centrality of a specific node x_i as the proportion of shortest paths that contain x_i , the flow-based centrality calculates the proportion of the flow that passes through x_i . The advantage of the new measure is two-fold. Firstly, the flow-based metric can be applied to a wider variety of network datasets, for it is defined for both binary and non-binary graphs. Secondly, the new metric considers all the independent paths between all pairs of nodes in the network, instead of merely the shortest paths, which is more realistic in some application scenarios. In the Linear Threshold Model of influence maximization, weights on the edges reflect the influence magnitude between

¹The betweenness of a node is the fraction of shortest paths between all possible pairs of nodes that pass through this node [12].

nodes, and the influence propagates through all the possible paths, which makes the flow-based centrality a promising alternative. However, the flow-based metric proposed precludes the possibility that nodes in the system may not be willing to forward the influence received. In this paper we generalize the original maximum flow problem to account for node noncooperativeness. To the best of our knowledge, this paper is the first to discuss the possibility of utilizing network flow theory to model node noncooperation in networks.

In the field of communication and computer networks, [13], [14] and [15] respectively studied the problem of noncooperative routing, load balancing, and flow control as examples of the impact of noncooperation on the overall performance of networked systems. While previous work focused on the inefficiency caused by node noncooperation in traditional transportation and communication networks, here we address this question in the online social advertising scenario. In this paper we mainly focus on the final influence size when the system reaches steady state. Therefore we assume a two-tiered and static differentiation of node cooperativeness in the system, i.e. seed nodes are cooperative to propagate the influence while the ordinary nodes are only partly willing to do so, and node cooperativeness will stay constant over time. We will leave the transient distribution of node cooperativeness (e.g., cooperativeness is related to incentives, social distance, and changes over time) for future investigation.

Issues of extracting influence diffusion model parameters from real network datasets are beyond the scope of the paper. Readers interested are suggested to refer to some machine learning papers, e.g. [16], [17], [18], and [19].

III. MODEL

In this section, we formulate the influence maximization problem and describe the two mainstream diffusion models.

A. Problem formulation

We consider an online social network (OSN) as a directed graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set of nodes (OSN users) and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges (social ties) in the network. We also denote by $\mathcal{N}_u \subseteq \mathcal{V}$ the set of neighbors of node u . Each node in the system can either be active or inactive. As more neighbors of an inactive node become active, the more likely it will switch to active. A node cannot return to the inactive state once it becomes active. All nodes are inactive at the beginning of the influence propagation process and marketing practitioners try to initially activate K nodes to seed the information cascade in the social network. The process ends when no more nodes can be activated. The influence maximization problem is defined as follows: Determine the K -node set to achieve the maximal expected active nodes at the end of the process.

B. Diffusion models

In the literature there are mainly two diffusion models utilized to mimic the propagation effect, namely, the Linear Threshold Model (LTM) [20] and the Independent Cascade

Model (ICM) [21]. In LTM, a node i has a weight of influence $b_{i,j}$ on node j . Node j is pre-assigned a uniformly distributed threshold $\theta_j \in [0, 1]$. At each discrete time step, node j is considered activated when the sum of weights from its active neighbors exceeds the threshold θ_j . In ICM, a node i activated at time t has a probability $p_{i,j}$ to successfully activate its inactive neighbor j at time $t + 1$. Node i does not have any further opportunities to activate j again whether it succeeds or not. In this paper we mainly focus on the study of LTM and defer the ICM case as future work.

IV. NONCOOPERATIVE INFLUENCE MAXIMIZATION PROBLEM

We study the inefficiency of influence maximization in terms of the final influence size because of node noncooperation in this section. We first introduce a noncooperative linear threshold model, and then provide some useful properties of the model. After that we describe an analytical model based on the generalized maximum flow problem which accounts for node noncooperativeness, and then we propose a maximum-flow-based centrality measure and illustrate it under the example of the newly proposed cluster-based seed node selection heuristic [7].

A. A noncooperative linear threshold model

An implicit assumption under LTM is that nodes in the system are all cooperative to propagate the influence, which we think may be unrealistic as discussed in Section I. In general, each node can have distinct cooperativeness levels compared to other nodes. For the ease of simulation and without loss of generality here we assume a two-tiered and static differentiation of node cooperativeness in the system, i.e. seed nodes are cooperative to propagate the influence while the ordinary nodes are only partly willing to do so. We further define Node i to be α -cooperative so that weights of edges coming out of Node i are multiplied by a factor $\alpha \in [0, 1]$.

Formally, under the generalized noncooperative LTM, the activation criteria for Node j becomes

$$\theta_j \leq \sum_{i \text{ active neighbor of } j} \alpha_{i,j} b_{i,j},$$

where $\alpha_{i,j} \in [0, 1]$ measures the cooperativeness of Node i on its neighbor Node j . Under the two-tiered and static differentiation of node cooperativeness, we can set $\alpha_{i,j} = 1$ if Node i belongs to the seed-node set because they are cooperative to contribute all their influence capacity that can be imposed on their neighbors and $\alpha_{i,j} = \alpha < 1$ otherwise, that is all non-seed nodes are considered α -cooperative, which means that they will reserve some of their channel capacities due to various reasons.

B. Properties of the model

We now discuss some nice properties of the noncooperative LTM. First we define a set function $\sigma(\cdot)$ to be submodular if $\sigma(S \cup \{v\}) - \sigma(S) \leq \sigma(T \cup \{v\}) - \sigma(T)$ for all $v \in \mathcal{V} \setminus T$ and $S \subseteq T$, i.e., $\sigma(\cdot)$ satisfies a “diminishing returns” requirement:

the marginal gain from adding a node to a set T is at most the same as the marginal gain from adding the same node to a subset of T . In addition, we say that $\sigma(\cdot)$ is monotone if $\sigma(T) \leq \sigma(S)$ for all $S \subseteq T$, that is, $\sigma(\cdot)$ will at least stay the same after adding elements to the original set. We also define a greedy algorithm as follows: starting from an empty set, the algorithm iteratively selects a seed which achieves the highest incremental change of $\sigma(\cdot)$. [22] provides a result that a non-negative, monotone submodular objective function can be approximated to within a factor of $(1 - 1/e)$ (around 63%, here e is the base of the natural logarithm) using greedy algorithm, which is as follows:

Theorem 1. [22] *The greedy algorithm is a $(1 - 1/e)$ approximation for a non-negative, monotone submodular objective function.*

[6] further proves that the greedy algorithm can also achieve $(1 - 1/e)$ approximation for the influence maximization problem by proving that the final influence function $\sigma(\cdot)$, which is the expected number of the final active node in the network at the end of the diffusion process, is submodular. Based on [6], we prove that under the two-tiered and static differentiation of node cooperativeness, the influence function under the proposed noncooperative LTM also satisfies the requirement of submodularity, so that a greedy algorithm can also achieve the same $(1 - 1/e)$ performance guarantee.

Lemma 1. [6] *The influence function $\sigma(\cdot)$ is submodular for an arbitrary instance of the LTM.*

Theorem 2. *Under the two-tiered and static differentiation of node cooperativeness, the influence function $\sigma(\cdot)$ of noncooperative LTM is submodular.*

Proof: Since the cooperativeness parameter $\alpha_{i,j}$ can be considered constant (α or 1) under the two-tiered and static cooperativeness differentiation, the noncooperative LTM is equivalent to a standard LTM in which $b'_{i,j} = \alpha_{i,j} b_{i,j}$. Thus according to Lemma 1, the influence function of noncooperative LTM is also submodular. ■

Proving that the influence function under noncooperative LTM also satisfies the requirement of submodularity not only shows that the model has a performance guarantee, but also implies that the incentive needed for the advertising campaign should show similar property, since the amount of incentive needed is closely related to the initial active set size. It is also intuitively satisfying that incentive as a function of initial active users would show a “diminishing returns” property. The detailed study of the relationship between the amount of incentive and seed-node set size in noncooperative influence maximization problem will be our future work.

Note that we have to use Monte-Carlo simulations to estimate $\sigma(\cdot)$ because there is no explicit formula for the influence function. This means that we can obtain a $(1 - 1/e - \epsilon)$ approximation with small ϵ if we run a large number of simulations.

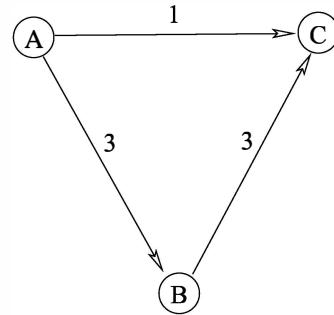


Fig. 1. Maximum flow from Node A to Node C

C. The framework

In LTM, it is natural to consider edges of the graph as channels of influence propagation between pairs of people, and the weight of each edge as the *capacity* of the channel linking them. The capacity is the maximum amount of information that can be transmitted between them. It is reasonable to assume that a pair of people in the social network who are socially close are connected by a channel with large capacity while pairs less close are assigned “narrower” channels, i.e., with lower capacity.

In influence maximization problems, the key is to measure the influence Node A can impose on Node B quantitatively. The characteristics of online social advertising, such as random communication and influence intentionally channeled through multiple intermediaries renders information, or influence flow to take more than one route. Thus a reasonable metric is the overall information flow along all possible paths that connect them, for it reflects the total amount of influence that can be transmitted between two nodes. In other words, the flow between two nodes not only depends on the capacity of the direct channel, but also on the capacities of all channels on all the indirect paths that connect the two. For example, to measure the total influence Node A can impose on Node C in Figure 1, it is incomplete to only count the direct link between A and C. The path passing Node B as intermediary should also be considered.

Ford and Fulkerson [23] formulated a problem to calculate this kind of network flow, namely, the maximum flow problem. The problem defines a flow from a source to a sink constrained by two conditions: (1) the flow value cannot exceed the channel capacity, (2) the flow into any intermediate node must be equal to the flow out of that node. Under this formulation, in Figure 1, the maximum flow between Nodes A and C would be $1(A \rightarrow C) + 3(A \rightarrow B \rightarrow C) = 4$, which is the maximal possible amount of flow that can be transmitted from Node A to Node C.

However, the second constraint, which is often referred to as *conservation of flows* implicitly assumes that an intermediate node will automatically forward all information it receives from its neighbors. That is all nodes in the network are assumed to be cooperative, which is unreasonable as discussed in Section I. So we generalize the traditional maximum

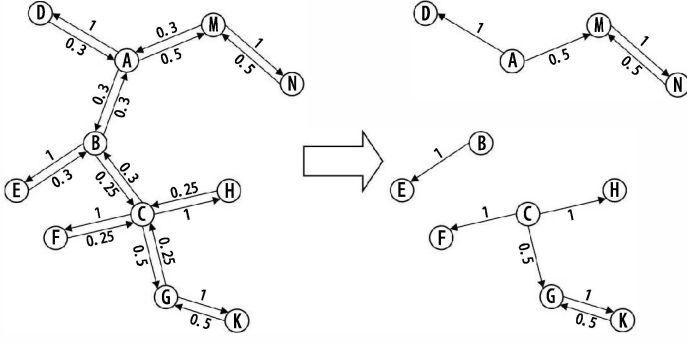


Fig. 2. Clustering with $\omega = 0.4$, ω is the cut-off threshold utilized in cluster identification scheme as described in [7]

flow problem by introducing a single variable representing node noncooperativeness. The mathematical formulation of the generalized maximum flow problem is as follows:

Let $s, t \in \mathcal{V}$ be the source and sink of the network \mathcal{G} . The capacity of an edge is a mapping $c : \mathcal{E} \rightarrow \mathbb{R}^+$, denoted by c_{uv} . It represents the maximum amount of flow that can pass through an edge. A flow is a mapping $f : \mathcal{E} \rightarrow \mathbb{R}^+$, denoted by f_{uv} , subject to the following two constraints:

- 1) $f_{uv} \leq c_{uv}$, for each $(u, v) \in \mathcal{E}$.
- 2)

$$\sum_{u:(u,v) \in \mathcal{E}} f_{uv} - \sum_{u:(v,u) \in \mathcal{E}} f_{vu} = b_v, \quad (1)$$

for each $v \in \mathcal{V} \setminus \{s, t\}$. $b_v \geq 0$ is the difference between the sum of flow coming in and out of Node v . This is the amount of flow Node v chooses to reserve, so it naturally measures the noncooperativeness of Node v , i.e. the bigger b_v is, the more noncooperative Node v is. Note that $b_v = 0$ means that Node v is willing to forward all its incoming influence, and the corresponding problem becomes the traditional maximum flow problem.

The objective of the proposed generalized maximum flow problem is to maximize $|f_s| = \sum_{v:(v,t) \in \mathcal{E}} f_{vt}$. $|f_s|$ will be utilized to calculate the centrality metric in the following seed node selection heuristic.

D. Maximum-flow-based centrality measure

We first give a brief description of the flow-based centrality value proposed in Section IV-C under the example of the clustering scheme proposed in [7]. A detailed discussion of the theoretical rationale behind the cluster-based scheme and parameter choosing can be found in [7].

In social networks, nodes considered influential (i.e. with high betweenness value) tend to form tightly-knit regions, which is known as homophily [24]. Thus in traditional betweenness-based scheme, selecting seed nodes merely according to their betweenness value will inevitably cause overlapping activations, which is wasteful as multiple initial nodes are expended to cover a small region. The cluster-based heuristic can avoid this problem by deliberately diversifying seed nodes into different regions (i.e. clusters) of the network instead of only searching intensively in a small area. This

motivates us to illustrate the maximum-flow-based centrality measure under the example of the cluster-based algorithm. Also calculating this centrality measure in clusters instead of the whole graph can largely reduce the computational cost. Studying the impact of other clustering strategies (e.g. [25]) on our selection scheme will be our future work.

To select the target node, we need to find the most influential node in order to maximize the number of activated nodes. [7] defined a weighted degree metric which is the sum of weights from a node to all its neighbors. It is obvious that the proposed metric confines a node's influence only in its own social circle while it is more reasonable to measure a node's importance in terms of its overall influence in its own cluster, or even in the whole graph. Moreover, considering the influence in only one hop makes it impossible to take node noncooperativeness into consideration. For example, selecting a seed node merely according to its weighted degree value would possibly result in choosing an "influential" node with highly noncooperative neighbors, then the influence will be confined in its own social circle and the performance would be unsatisfactory. In this case it may be better to choose a less "influential" node but with more cooperative neighbors, who are willing to further propagate the influence they receive.

We define an alternative flow-based centrality metric γ_i which is the sum of maximum flow values from Node i to Node j , where j is a node in the same cluster, and the flow values are calculated one after another. The cluster head is the node with biggest γ_i . For example, in Figure 2 we will select Node A as the cluster head of the corresponding cluster because $\gamma_A = 1(A \rightarrow D) + 0.5(A \rightarrow M) + 0.5(A \rightarrow N) = 2$ (suppose all nodes in the network are cooperative, i.e. $b_v = 0$ in (1), for each $v \in \mathcal{V}$), which is the highest in its cluster. Nodes B and C will be selected as cluster heads under similar calculations.

Finally, a formal statement of the flow-based seed-node selection method under the example of cluster-based heuristic is given in Algorithm 1.

V. EVALUATION

In this section, we study the effect of node noncooperation on the system performance in terms of the final active set size. We test our proposed algorithm on real large academic collaboration networks.

A. Dataset and influence model

We evaluated our proposed heuristic on Arxiv's co-authorship network under the General Relativity and Quantum Cosmology category [26]. The graph constructed contains 4158 nodes and 26850 edges. Each node is an author, and an edge between two authors i and j means that they have co-authored a paper. We consider the co-authoring relationships between two authors only once in case that two authors have co-written more than one paper.

To determine the weights of edges in LTM, we further define that if Node i with degree d_i connects to Node j , edge (j, i) has weight $\frac{1}{d_i}$. An intuitive explanation for this assignment

Algorithm 1 Cluster-based heuristic

Let e_1, \dots, e_N be nodes and C_1, \dots, C_M be clusters

Input:

Network $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with weight $b_{i,j}$ for $(i, j) \subseteq \mathcal{E}$.

A given integer K and a pre-determined threshold ω ;

Output:

The final target set S_t ;

- 1: $S_t \leftarrow \emptyset$
- 2: **if** $(i, j) \subseteq \mathcal{E}$ and $b_{i,j} < \omega$ **then**
- 3: delete (i, j) from \mathcal{E} .
- 4: **end if**
- 5: Use DFS to identify M clusters from \mathcal{G} .
- 6: **if** $M < K$ **then**
- 7: set a larger ω .
- 8: **GOTO** 1.
- 9: **end if**
- 10: Choose the K largest clusters $\{C_1, \dots, C_K\}$.
- 11: **for** $i = 1$ to K **do**
- 12: find node e with the largest value of γ in C_i .
- 13: $S_t \leftarrow S_t \cup \{e\}$
- 14: **end for**

is that the more friends a node has, the less likely it will be influenced by a single friend.

B. Cooperativeness differentiation

In order to differentiate node cooperativeness level in the simulation under noncooperative LTM, for the flow-based centrality value calculation, we can take nodes' noncooperativeness into consideration by assuming that intermediate ordinary nodes will reserve $1 - \alpha$ of its incoming flow, i.e. they will only forward α of the flow it has received. Then (1) becomes

$$\sum_{u:(u,v) \in \mathcal{E}} f_{uv} - \sum_{u:(v,u) \in \mathcal{E}} f_{vu} = (1 - \alpha) \sum_{u:(u,v) \in \mathcal{E}} f_{uv}, \quad (2)$$

for each $v \in \mathcal{V} \setminus \{s, t\}$.

We compare the performance of seed node selection algorithms in terms of the final active size with $\alpha = 80\%$, $\alpha = 50\%$ and $\alpha = 20\%$ for all ordinary nodes, respectively. For the cluster identification scheme, we set the cut-off threshold $\omega = 0.4$.

C. Result

Here we study the effect of node cooperativeness level on the performance of seed node selection strategy. The result is shown in Figure 3. The results are obtained as averages of 500 simulation runs. The x-axis represents the cooperativeness level α and the y-axis represents the final active set size. From the figure we can see that the performance of the seed node selection scheme improves as α increases, and the performance of algorithms with a larger target set size is always more satisfactory. An intuitive implication for online social advertising practitioners is that in order to achieve an

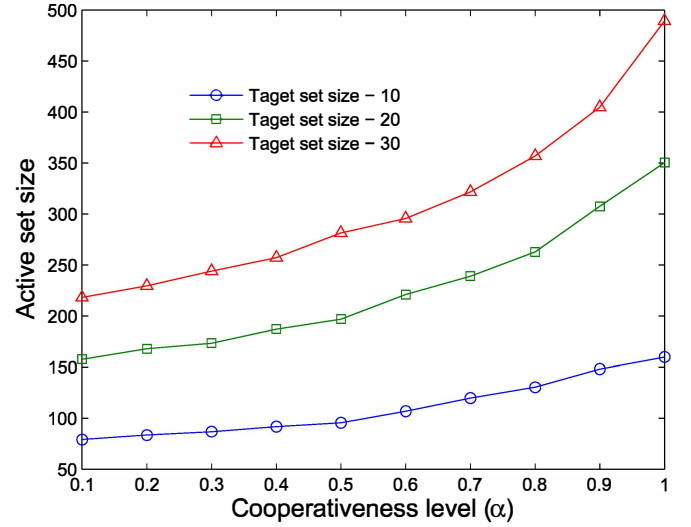


Fig. 3. Performance under LTM for different α and target set size

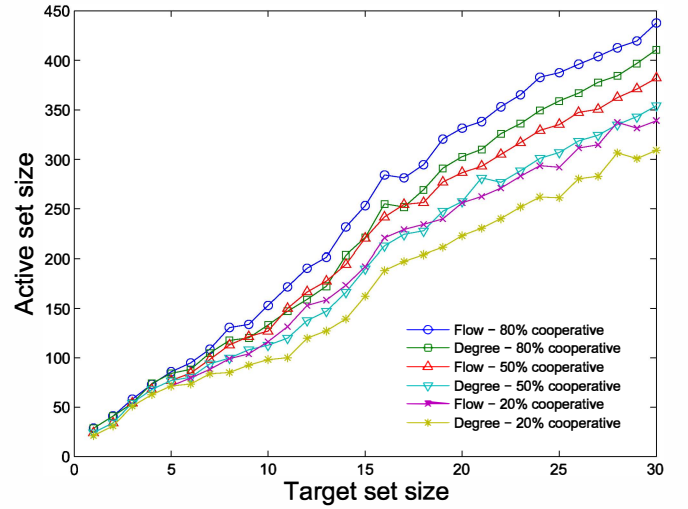


Fig. 4. Comparison of performance between fully and partially cooperative nodes

effective online viral marketing campaign, they should offer enough incentive to recruit enough initial product adopters to seed the influence cascade. They should also provide incentive to other nodes in the system so that all nodes in the network would be cooperative to propagate the influence.

Figure 4 shows the performance comparison of cluster-based heuristic with the weighted degree centrality metric and our proposed flow-based centrality metric when $\alpha = 80\%$, $\alpha = 50\%$ and $\alpha = 20\%$ (ordinary nodes are only α -cooperative as described in Section V-B). The x-axis represents the number of initially active nodes and the y-axis represents the final active set size. We can see that if we compare the performance of the two metrics separately, the average active set size is bigger when α is higher. This further verifies our

conclusion in Figure 3 that if nodes in the system choose to reserve its resource (i.e. channel capacity), the influence propagation process cannot achieve a satisfactory performance. Thus it is important to provide incentives for nodes in the system to be cooperative.

Figure 4 also strongly supports our proposition that the new flow-based metric is more robust against node noncooperation in the system. From the figure we can see that not only the final active set size under the flow-based metric outperforms the weighted degree metric with the same α , but also the final active size when nodes are 50% cooperative is comparable to the final active set size under the weighted degree metric when nodes are 80% cooperative. Moreover, the final active set size under the flow-based metric when nodes are only 20% cooperative is even better than weighted degree metric when nodes in the system are 50% cooperative in most of the target set size. The key implication is that our proposed metric is probably the best choice to achieve satisfactory performance for marketing practitioners with restricted budget to encourage nodes to be cooperative.

VI. CONCLUSION

In this paper, we study the influence maximization problem in noncooperative social networks. We generalize LTM to take node noncooperation into consideration and provide a provable approximation guarantees for the noncooperative influence maximization problem. Then we propose an analytical model based on the generalized maximum flow problem which captures a node's behavior in maximizing influence. Based on this, we develop a new seed node selection strategy for LTM which accounts for user noncooperativeness. Extensive simulations on large collaboration networks show that our proposed strategy outperforms the original scheme under various noncooperative scenarios, that is, the proposed metric is more robust when nodes are not cooperative. The evaluation also shows the importance of cooperation and incentive in maximizing influence. In this paper we assume a two-tiered and static differentiation of nodes' cooperativeness in the system, i.e., seed nodes are cooperative to propagate the influence while the ordinary nodes are only partly willing to do so. In the future, we would like to study the impact of transient distribution of node cooperativeness (e.g., cooperativeness is related to incentives, social distance, and changes over time) on the performance of influence maximization algorithms. Another potential future direction is to study the impact of noncooperation in other influence diffusion models, including those which do not satisfy the submodularity requirement. Of course, generalizing the noncooperative LTM to ICM and finding a remedy to the inefficiency caused by node noncooperation in ICM are interesting research directions.

REFERENCES

- [1] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- [2] J. E. Phelps, R. Lewis, L. Mobilio, D. Perry, and N. Raman, "Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email," *Journal of Advertising Research*, vol. 44, no. 04, pp. 333–348, 2004.
- [3] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Trans. Web*, vol. 1, no. 1, 2007.
- [4] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 57–66.
- [5] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2002, pp. 61–70.
- [6] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137–146.
- [7] X. Fan and V. O. K. Li, "The probabilistic maximum coverage problem in social networks," in *Proceedings of the IEEE GLOBECOM '11*, 2011.
- [8] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.
- [9] A. Post, V. Shah, and A. Mislove, "Bazaar: strengthening user reputations in online marketplaces," in *Proceedings of the 8th USENIX conference on Networked systems design and implementation*, 2011.
- [10] Z. Liu, H. Hu, Y. Liu, K. W. Ross, Y. Wang, and M. Mobius, "P2P trading in social networks: The value of staying connected," in *Proceedings of the IEEE INFOCOM '10*, 2010.
- [11] L. C. Freeman, S. P. Borgatti, and D. R. White, "Centrality in valued graphs: A measure of betweenness based on network flow," *Social Networks*, vol. 13, no. 2, pp. 141–154, 1991.
- [12] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
- [13] T. Roughgarden and É. Tardos, "How bad is selfish routing?" *J. ACM*, vol. 49, no. 2, pp. 236–259, 2002.
- [14] H. Kameda and O. Pourtallier, "Paradoxes in distributed decisions on optimal load balancing for networks of homogeneous computers," *J. ACM*, vol. 49, no. 3, pp. 407–433, Oct. 2002.
- [15] H. Kameda and E. Altman, "Inefficient noncooperation in networking games of common-pool resources," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1260–1268, Sep. 2008.
- [16] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 7–15.
- [17] A. Goyal, F. Bonchi, and L. V. Lakshmanan, "Learning influence probabilities in social networks," in *Proceedings of the third ACM international conference on Web search and data mining*, 2010, pp. 241–250.
- [18] K. Saito, M. Kimura, K. Ohara, and H. Motoda, "Selecting information diffusion models over social networks for behavioral analysis," in *Proceedings of the 2010 European conference on Machine learning and knowledge discovery in databases: Part III*, 2010, pp. 180–195.
- [19] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 807–816.
- [20] M. Granovetter, "Threshold models of collective behavior," *American Journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.
- [21] J. Goldenberg, B. Libai, and E. Muller, "Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata," *Academy of Marketing Science Review*, vol. 2001, no. 9, pp. 1–18, 2001.
- [22] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions," *Mathematical Programming*, vol. 14, pp. 265–294, 1978.
- [23] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canada Journal of Mathematics*, vol. 8, pp. 399–404, 1956.
- [24] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.
- [25] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, 2004.
- [26] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densefication and shrinking diameters," *ACM Trans. Knowl. Discov. Data*, vol. 1, 2007.